Understanding Systematics

Mateus F. Carneiro, BNL

Neutrino Interaction School

DUNE

Brookhaven National Laboratory

What are we covering here:

- Intro
- Statistical and Systematic uncertainties
- Modeling your errors
- Typical systematics uncertainties
- Estimating systematics
- Propagating your errors
- Model comparison and error reporting
- Outro

Outline

What are we covering here:

- Intro
- Statistical and Systematic uncertainties
- Modeling your errors
- Typical systematics uncertainties
- Estimating systematics
- Propagating your errors
- Model comparison and error reporting
- Outro

Disclaimer: a lot of this talk in based on previous classes done by **Dr. Cheryl Patrick** and **Dr. Ben Messerly**. I'm also mostly using MINERvA results as examples for historical reasons (and the fact it's a xsec measurement dedicated experiment).
Thank you all for the collaboration.

Introduction

Neutrino
Interaction
School

Do we need errors?

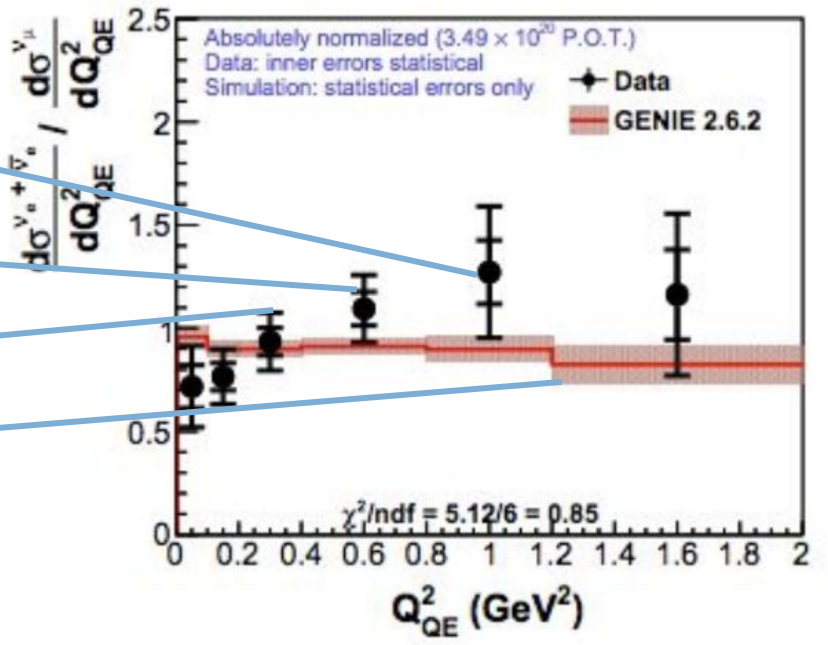
- Yes.
- Experimental Measurements are basically worthless without errors.
- Neutrino Interaction measurements main use is to validated models that describe the data better, and reducing other neutrino experiments systematic errors.
  - When we talk about uncertainties this circular arguments often show up.

Statistics Statistics Statistics

- Statistics is hard and complex.
- it's unfortunate that most of us don't have a solid base of probability and statistics and it's also impossible for me to give you that base just today (both because of competency and time).
- We use modern and sometimes specific techniques, textbooks may or may not be your friends here.
  - Backup slides with some literature references

Some quick terms

- Central Value (CV) is the "best guess" measurement.
- Statistical or systematic uncertainty depicted by inner bar.
- Total uncertainty depicted by outer bar.
- Uncertainty on model shown sometimes
- "Uncertainty" and "error" often used interchangeably.



Statistical and Systematic uncertainties

Neutrino Interaction School

Statistical vs Systematic

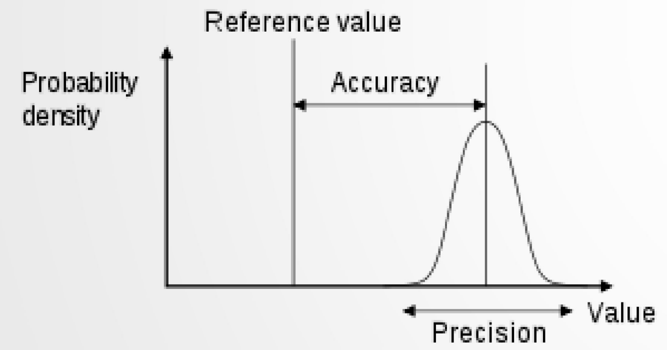
- A “**statistical uncertainty**” represents the scatter in a parameter estimation caused by fluctuations in the values of random variables. Typically this decreases in proportion to $1/\sqrt{N}$.

- A "**systematic uncertainty**" any error that's not a statistical error. More clearly, a systematic uncertainty is a possible unknown variation in a measurement, or in a quantity derived from a set of measurements, that does not randomly vary from data point to data point.”

- DO NOT TAKE THESE DEFINITIONS TOO SERIOUSLY. Not all statistical uncertainties decrease like $1/\sqrt{N}$. And more commonly, taking more data can decrease a systematic uncertainty as well, especially when the systematic affects different parts of the data in different ways.

From bias to uncertainty

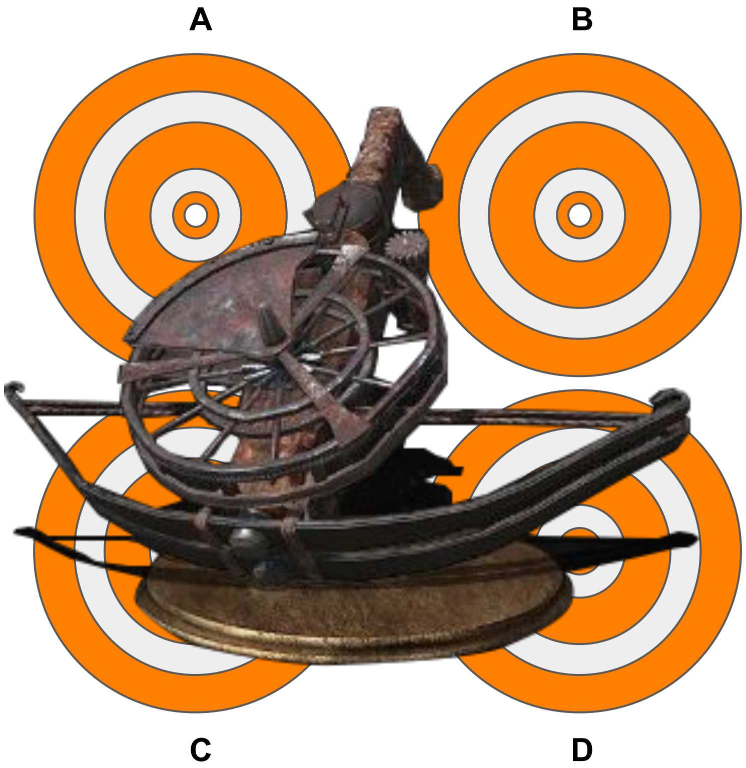
- **Precision** and **Accuracy**:
  - Precision is a description of random errors, a measure of statistical variability.
  - Accuracy describes systematic errors, that is, differences between the true and the measured value that are not probabilistic (or: bias).



- In particle physics, precision can be increased by accumulating more data
  - Equivalent to repeating the measurement

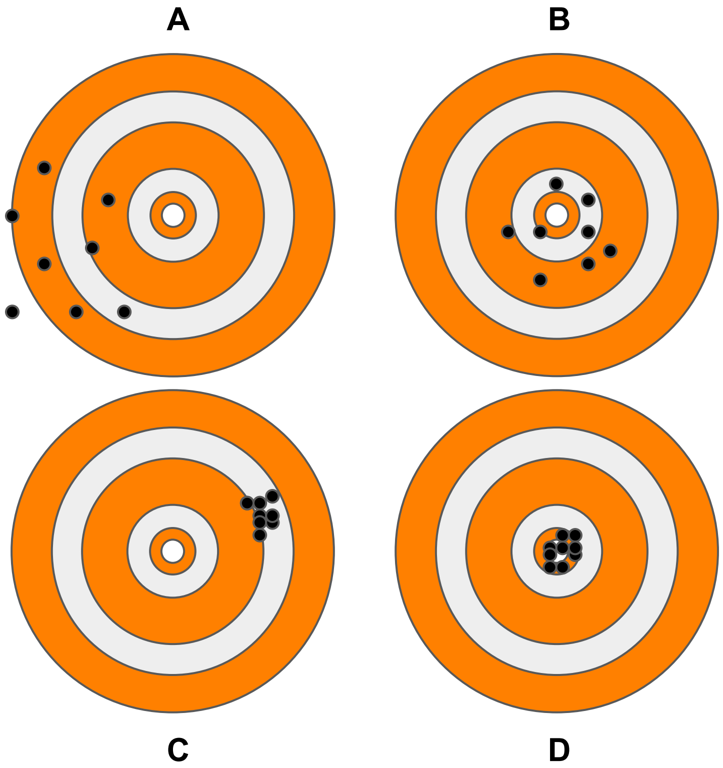
From bias to uncertainty

- Let us consider a toy setup:
  - One repeating dart machine aimed at the center of a target.
  - 4 example datasets:
    - A
    - B
    - C
    - D



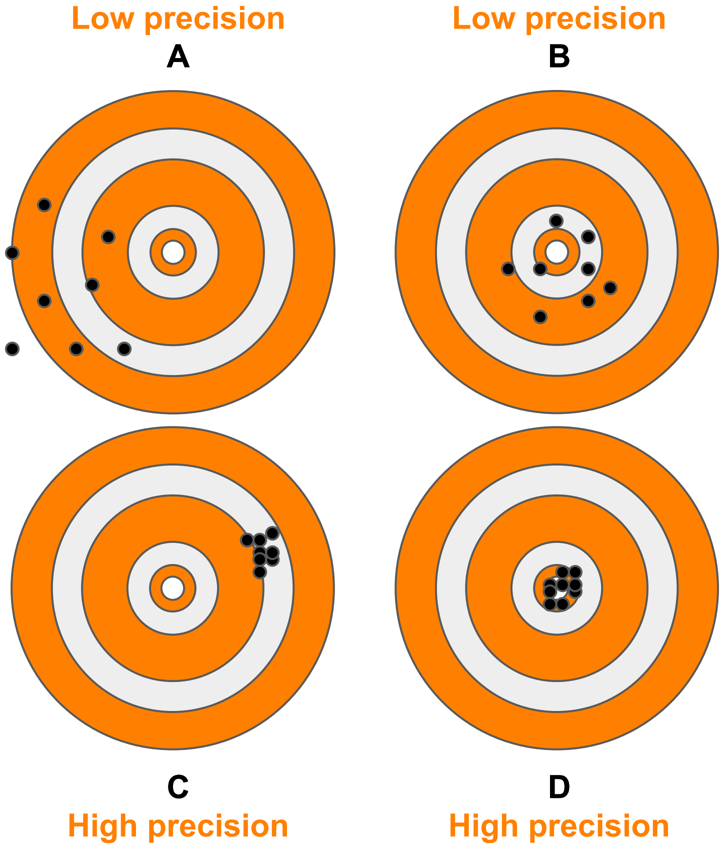
From bias to uncertainty

- Let us consider a toy setup:
  - One repeating dart machine aimed at the center of a target.
  - 4 example datasets:
    - A
    - B
    - C
    - D
- Can we infer the precision and accuracy in each case?



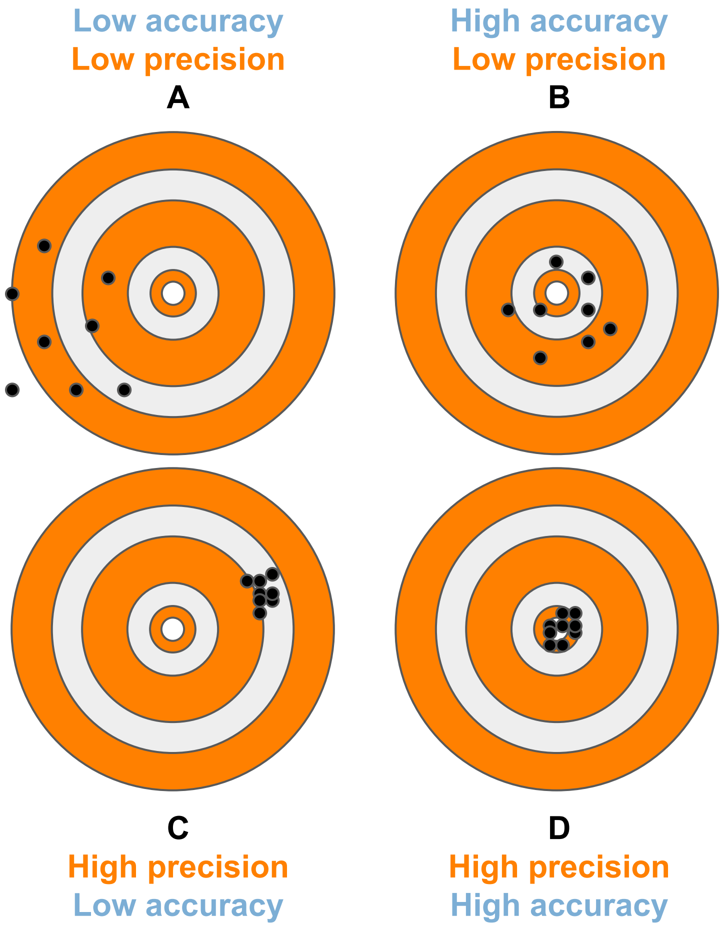
From bias to uncertainty

- 4 example datasets:
  - A: low precision
  - B: low precision
  - C: high precision
  - D: high precision
- C and D have **low statistical fluctuation** when several throws are made.
- A and B have a greater dispersion, or a **big statistical fluctuation**.



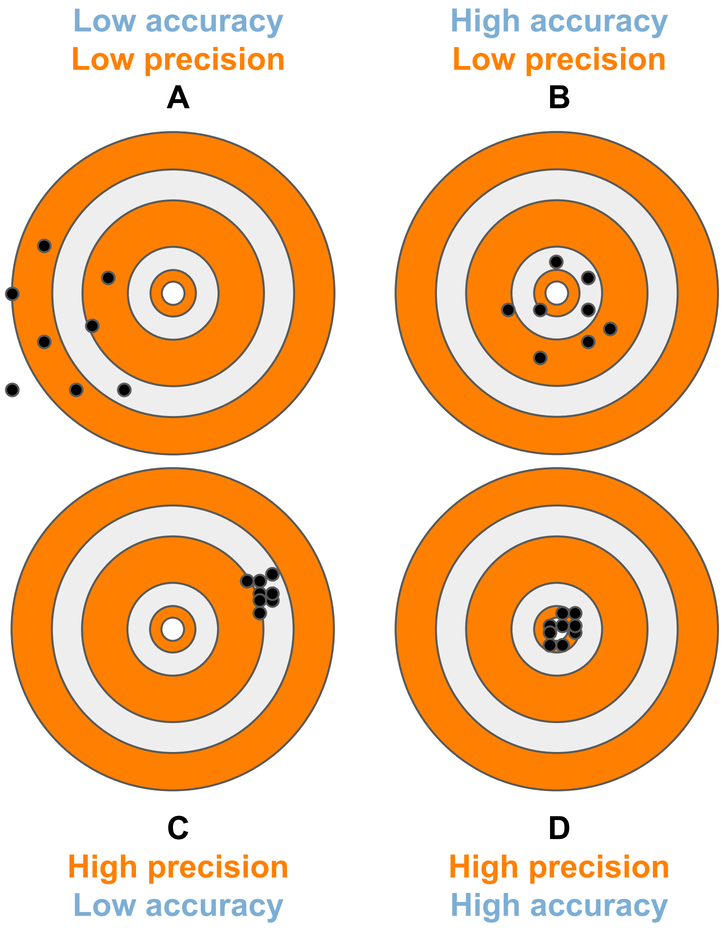
From bias to uncertainty

- 4 example datasets:
  - A: low accuracy
  - B: high accuracy
  - C: high accuracy
  - D: high accuracy
- B and D, since we are aiming at the center they are the ones getting it closer, with **no apparent systematic bias**.
- A and C miss the target by way more, there's some consistent error in the throws, or a **systematic bias**.



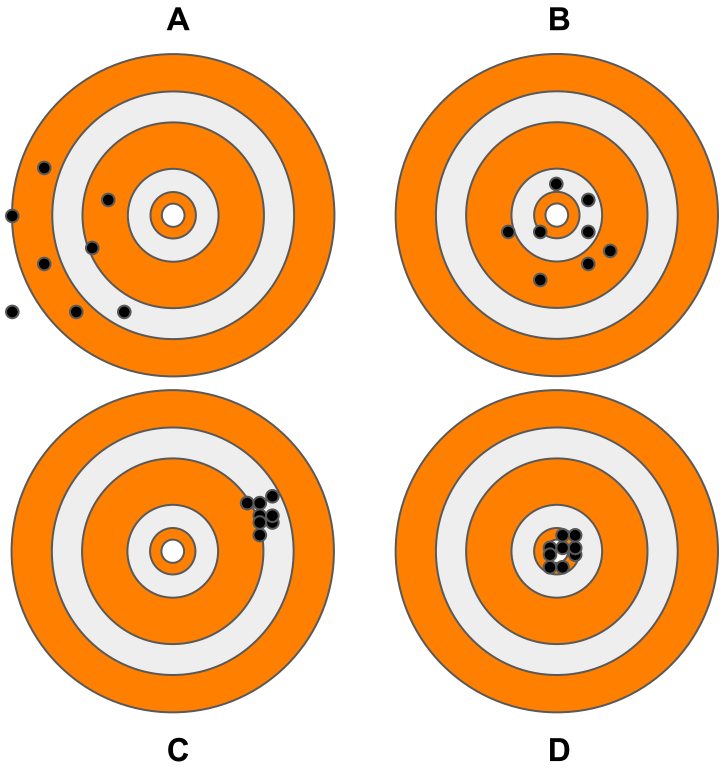
From bias to uncertainty

- 4 example datasets:
  - A: low precision - low accuracy
  - B: low precision - high accuracy
  - C: high precision - high accuracy
  - D: high precision - high accuracy
- Here one could just fix the machine to remove bias.
- Let's make this analogy more realistic.



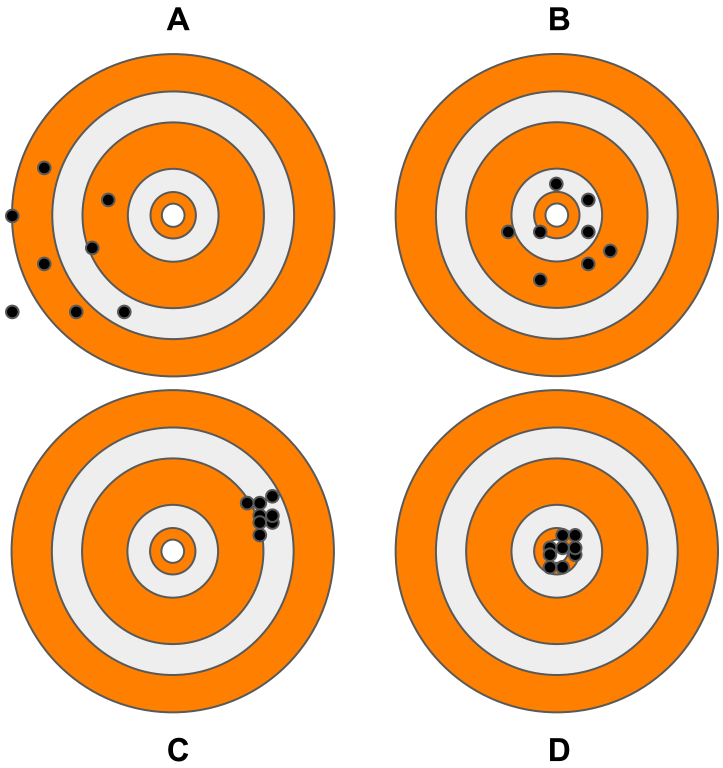
From bias to uncertainty

- Consider now the machine is aimed and you have no access to any of the setup.
- You have the target dataset, **where is the machine aimed** at and **how much can we trust this measurement**?



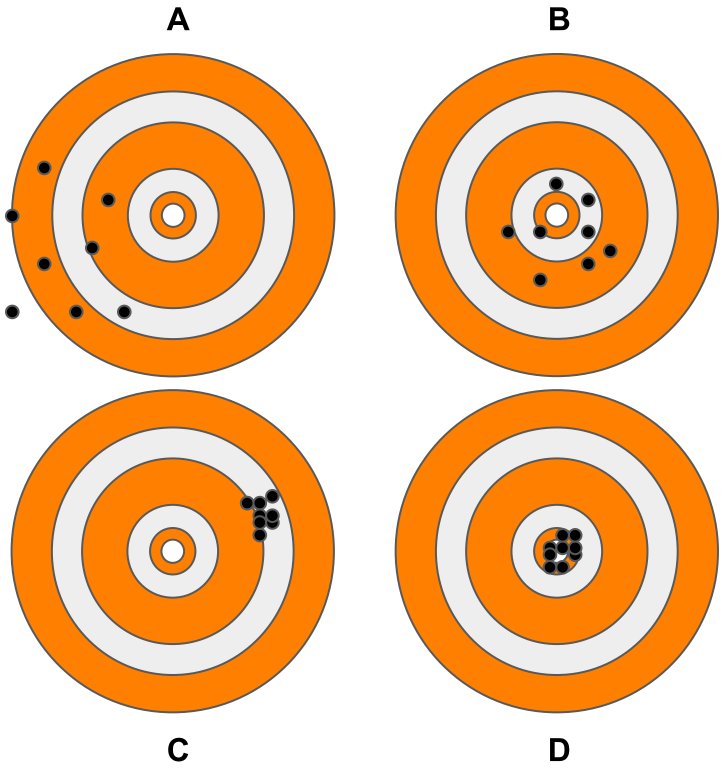
From bias to uncertainty

- The **statistical uncertainty** come from the amount of throws.
- Each new throw reduces the **statistical uncertainty** and increases your confidence in the central value as an average of all the positions.



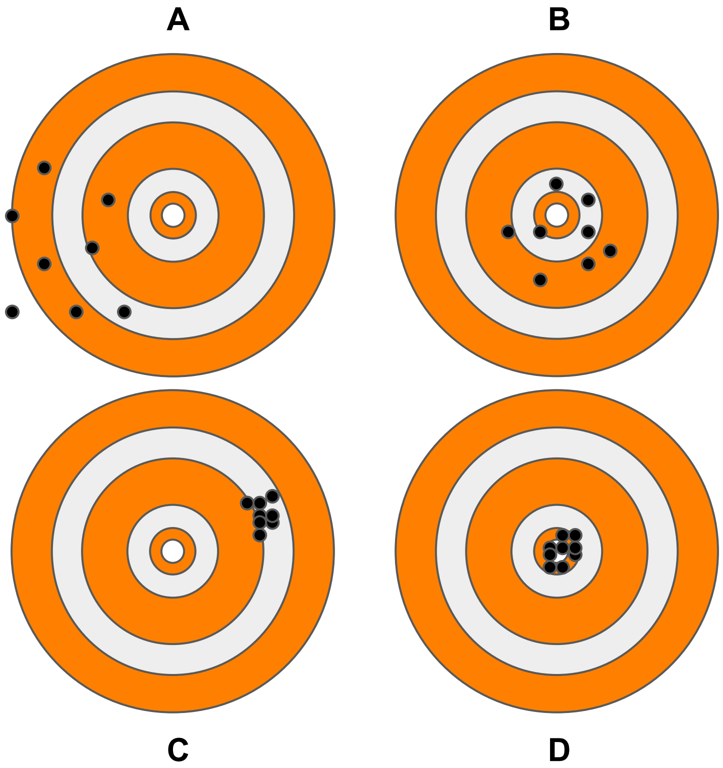
From bias to uncertainty

- Where do the systematic bias comes from tho?
  - You! Or rather from the conditions where that data was taken.
- You put the target here because you thought it was centralized, how are you measuring the distance in between the throws? Is there something in the environment pushing darts to one side? Maybe something in the way?
- It's hard to know, and we are focusing on quantifying the uncertainty, not fixing it.



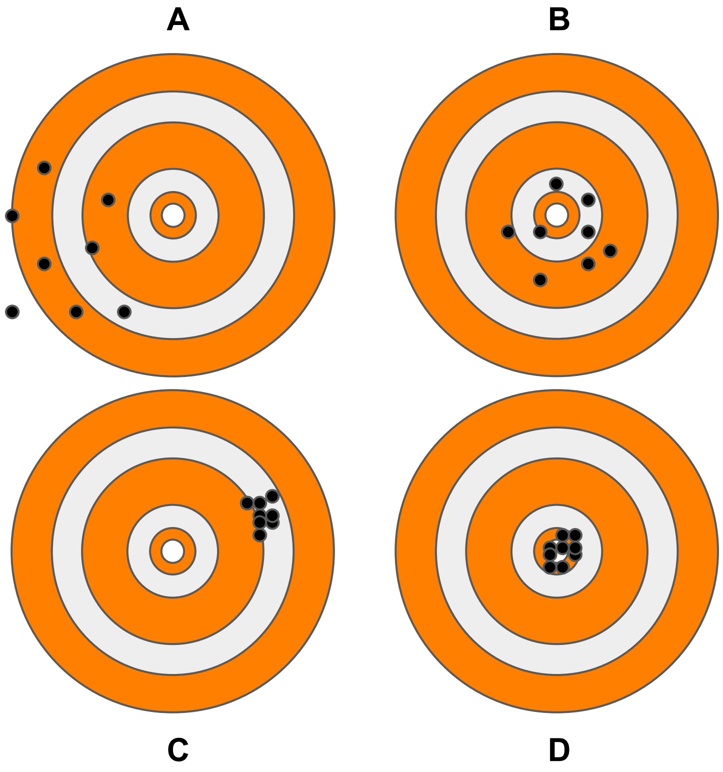
From bias to uncertainty

- What if you use all the knowledge you have of the machine and make a simulation?
- You can now compare your best guess with the actual data.
  - But, even if the overlap perfectly, you still need to quantify how much you can trust your simulation.



From bias to uncertainty

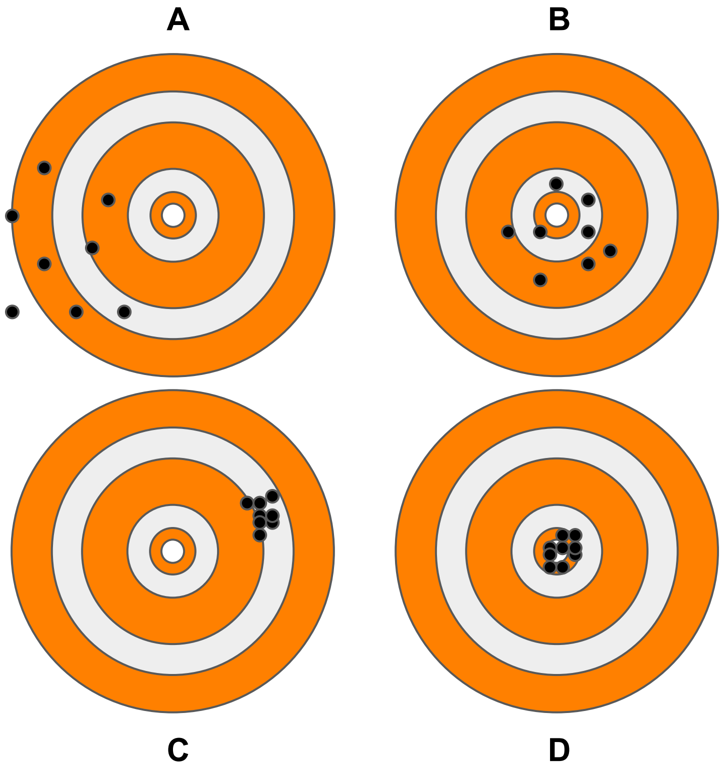
- Then make another simulation with something in the way and see how that affect the CV you had in the first try. Make another simulation considering you are the worst at measuring the position, another where… test all the assumptions you made in the first place.
- You can compare these changes to the CV and quantify how much you are sure of that CV.
- That's your systematic error in that measurement.



From bias to uncertainty

- Of course, this is a simplified case. None of these errors exist in a vacuum so you need to propagate them, some of our assumptions are based on physics models that we are also not fully sure of…

- In HEP data analysis **>90%** of the work is about thinking of clever ways to reduce or at least measure (and validate) systematics.

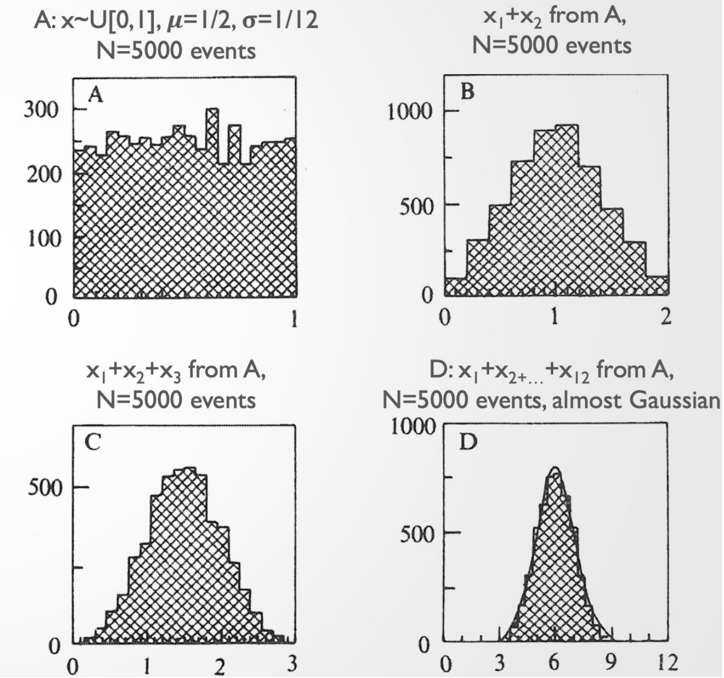


Modeling your error

Neutrino
Interaction
School

Probability distributions

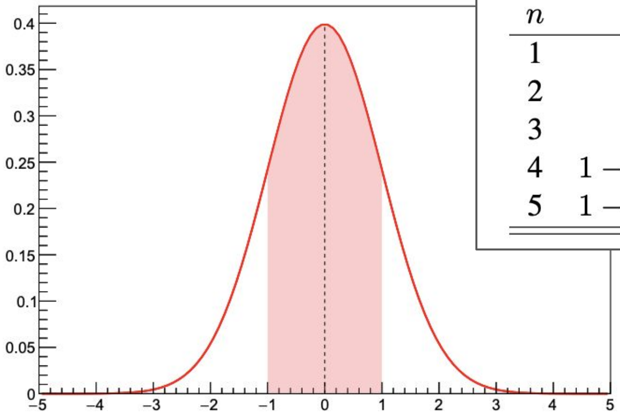
- Measurement uncertainties are often the sum of many independent contributions. We need a probability density function (PDF) to better quantify the underline probability.
- **Central limit theorem**:
  - When independent random variables are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.



NORMAL DISTRIBUTION (GAUSSIAN)

- A **Gaussian distribution** is given by:
  - where
    - μ - average value
    - σ - standard deviation of x
  - If $\mu = 0$ and $\sigma = 1$, a Gaussian distribution is also called **standard normal distribution.**
- Probability values corresponding to intervals $[\mu - n\sigma, \mu + n\sigma]$ for a Gaussian distribution are frequently used as reference.

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

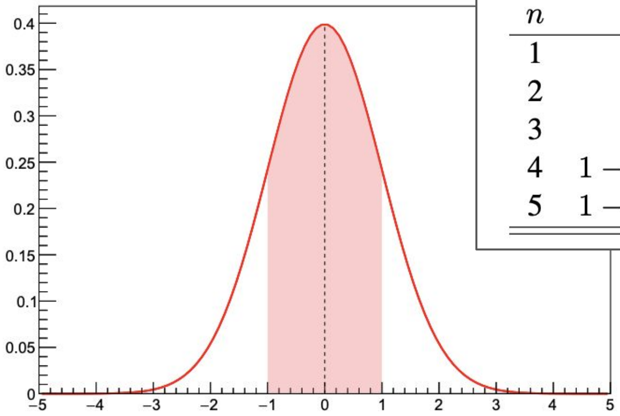


n	Prob.
1	0.683
2	0.954
3	0.997
4	$1 - 6.5 \times 10^{-5}$
5	$1 - 5.7 \times 10^{-7}$

NORMAL DISTRIBUTION (GAUSSIAN)

- Usually counting experiments (histograms!) are not described by Gaussian distributions but rather Poisson distributions.
- But for large counts, a Gaussian is a good approximation of a Poisson distribution.

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$



n	Prob.
1	0.683
2	0.954
3	0.997
4	$1 - 6.5 \times 10^{-5}$
5	$1 - 5.7 \times 10^{-7}$

(some)
Typical systematics uncertainties

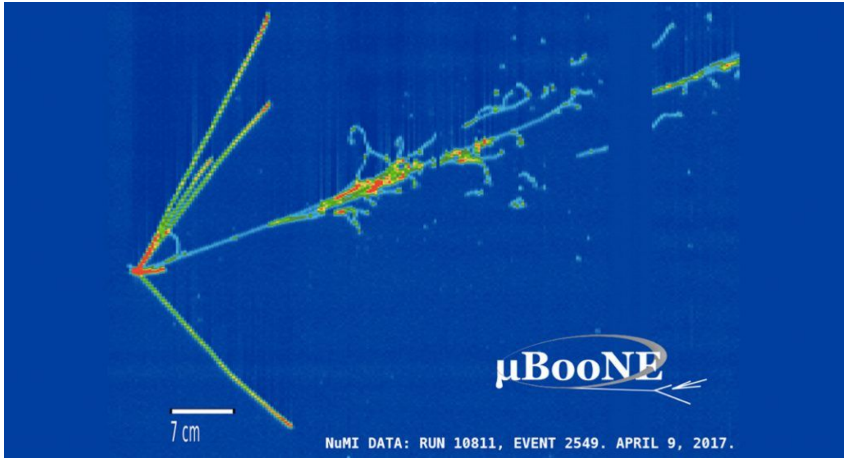
Neutrino Interaction School

Low level / detector calibration

- What's the cross talk on your photomultiplier?
- Which wires received the signal?
- How accurate is the timing?
- What's the alignment of the detector?

Event Reconstruction

- How accurately can we know the track angle?
- How well do we know the energy scale?
- What's the accuracy on the track position measured?
- Does the event have an overlapping cosmic ray?



Model Uncertainties

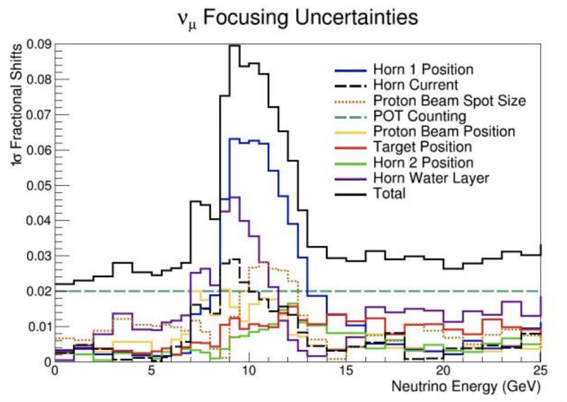
- There are many uncertainties in the models in our GENIE simulation
  - **Primary interaction rate** uncertainties
  - **Final-state interaction rate** uncertainties

- GENIE's parameter values and their uncertainties come from the results of previous experiments.

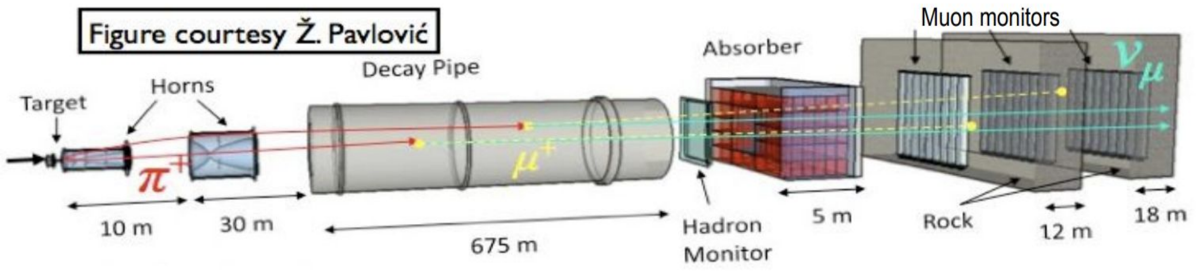
- GENIE sort of has you covered.
  - Uncertainties that GENIE doesn't consider might affect your analysis.
  - GENIE's uncertainty estimates might be not great.

Flux Uncertainties

- ❖ (some) Main uncertainties considered:
  - ➢ Beamline geometry
    - ■ Proton beam steering, size
    - ■ Magnetic horn positions, current
    - ■ Target position
  - ➢ Physics processes
    - ■ Probability of proton re-interacting in target
    - ■ Constrained by external data



Fermilab's NuMI Beam



Estimating systematics

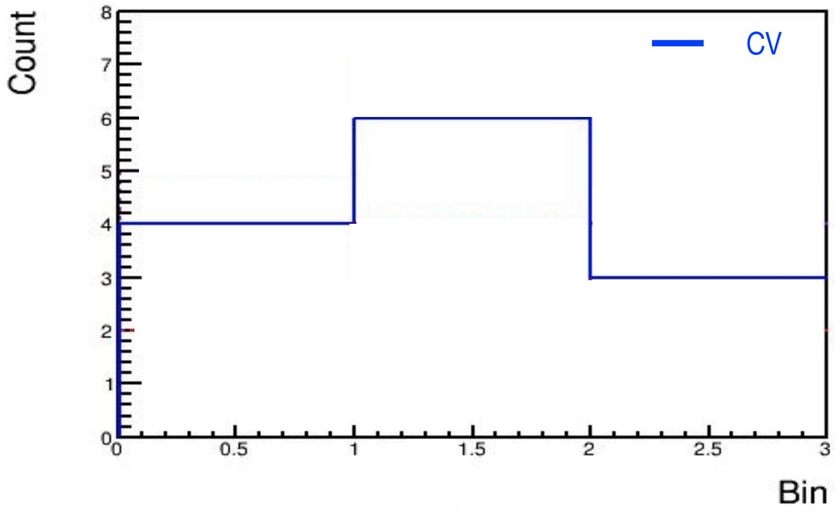
Neutrino
Interaction
School

How they affect the measurement?

- One-by-one we can determine how some uncertainty in the measurement process can leads to a small change.
- But how will these uncertainties affect our cross section?
  - Increasing the muon energy scale by 10% changes the reconstructed muon energy from 1GeV to 1.1GeV.
  - Shifting the vertex position by a distance between 1 and 10 cm might move the event in or out of the volume we’re studying.
  - Increasing pion production rate by 20% makes it any events with a pion 1.2 times as likely.

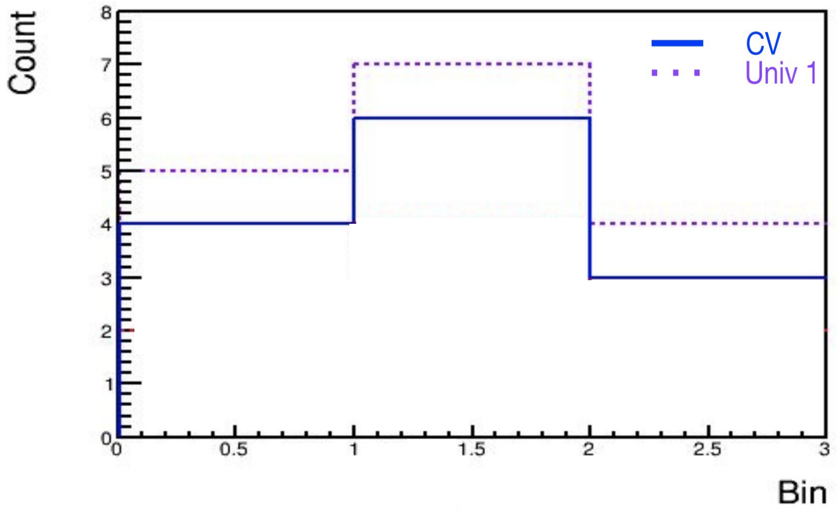
Alternate Universes

- Following the dart machine example:
  - 1. Make a best guess simulation, that's you Central Valeu (CV).



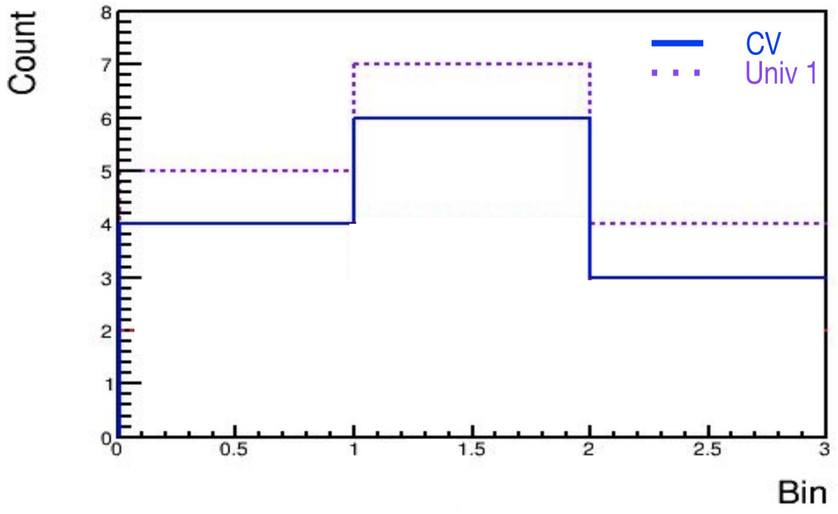
Alternate Universes

- Following the dart machine example:
  - 1. Make a best guess simulation, that's you Central Valeu (CV).
  - 2. Run the same simulation with a single parameter shifted by some amount (or best estimate of the 1σ uncertainty on that parameter).



Alternate Universes

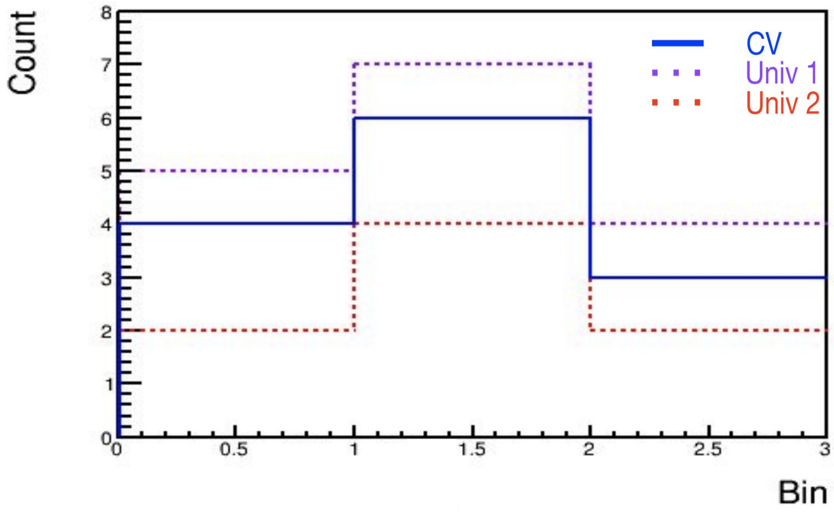
- In jargon, the 2nd simulation shows the CV in an "**alternate universe**" where, for example, pions are more likely to interact, or we measured all our vertex positions 10cm to the right.
- The difference between the distributions gives a measure of the uncertainty due to this parameter.



Alternate Universes

- We can actually keep trying new shifts of the same parameter.
- In this case, the uncertainty is the average of the differences between each universe and the central value.
- E.g.:
  $|N_1 - N_{CV}| = |7\text{-}6| = 1$
  $|N_2 - N_{CV}| = |4\text{-}6| = 2$

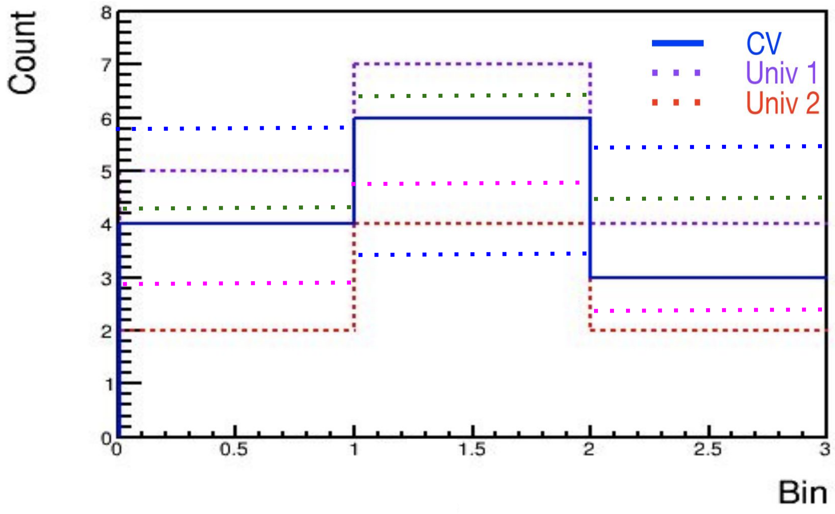
  Uncertainty is average: (1+2)/2 = 1.5
  Fractional uncertainty = 1.5/6 = 25%



(many) Alternate Universes

- We can actually keep trying new shifts of the same parameter.
- In this case, the uncertainty is the average of the differences between each universe and the central value.
- E.g.:
  $|N_1 - N_{CV}| = |7\text{-}6| = 1$
  $|N_2 - N_{CV}| = |4\text{-}6| = 2$

  Uncertainty is average: (1+2)/2 = 1.5
  Fractional uncertainty = 1.5/6 = 25%



(many) Alternate Universes

- There are generally 3 ways to make this variations:
  - Smearing and scaling of observables
  - Reweight techniques
  - Alternative simulations
- We study the most efficient way to make the variations and evaluate all systematic sources we consider.

Shifting the position of an event vertex in the detector will affect which events are in the detector's fiducial volume

Reweighting the probability that a CCQE event can occur will change the event count

Producing a whole new simulation sample with detector parameters shifted can affect the sample in many ways

Propagating errors

Neutrino
Interaction
School

Propagating errors

Errors in the Cross Section calculation

- ❖ Finally we are talking about xsections!
- ❖ If you managed to quantify your errors we should be ready to carefully consider them in the cross section calculation.
- ❖ Which terms have which kind of uncertainty?

Cross Section Formula

$$\underbrace{\boxed{\left(\frac{d\sigma}{dx}\right)_\alpha}}_{\text{Cross Section in bin alpha}} = \frac{\Sigma_j \underbrace{\boxed{U_{j\alpha}}}_{\text{Unfolding Matrix}} \left(\underbrace{\boxed{N_{data,j}}}_{\text{Selected Events}} - \underbrace{\boxed{N^{bkgd}_{data,j}}}_{\text{Subtract Backgrounds}}\right)}{\underbrace{\boxed{E_\alpha}}_{\text{Efficiency in bin alpha}} \underbrace{\boxed{(\Phi T)}}_{\text{Flux times the number of scattering centers}} \underbrace{\boxed{(\Delta x)}}_{\text{Bin Width Normalize}}}$$

- j represents the reconstructed bin
- Alpha represents the true bin
- x is the quantity you want to measure your cross section with respect to.

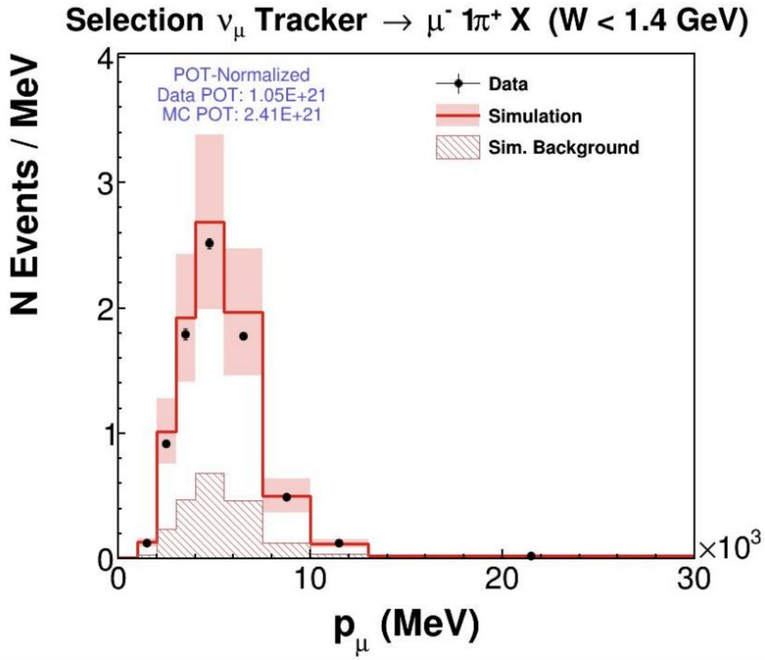
14 June 2021 Anne Norrick | So you want to measure a cross section... Fermilab DUNE Neutrino Interaction School 2021 47

Selected Events

$$\left(\frac{d\sigma}{dx}\right)_\alpha = \frac{\Sigma_j U_{j\alpha}(\boxed{N_{data,j}} - N^{bkgd}_{data,j})}{E_\alpha(\Phi T)(\Delta x)}$$

- Event selection
  - Selected data events – only have statistical uncertainty bars.
  - Selected MC events –statistical and systematic uncertainty bars.

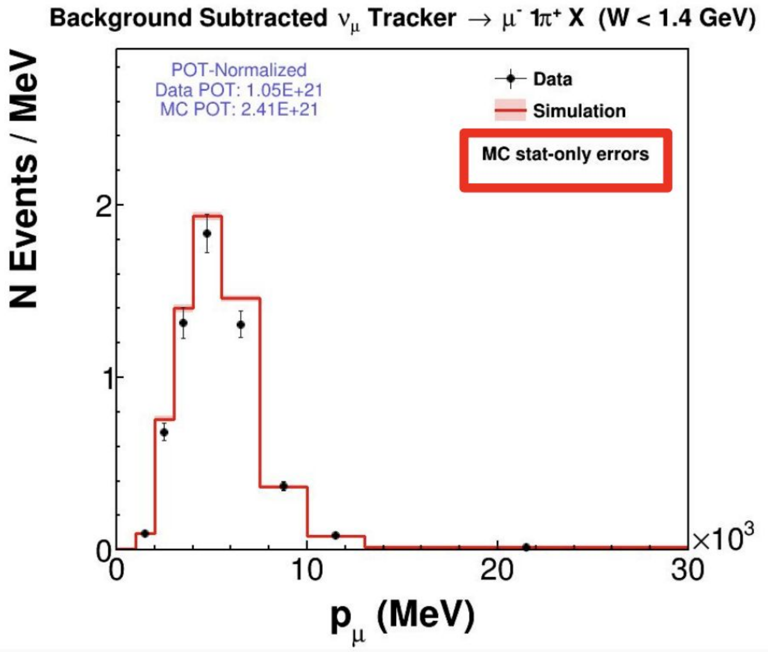
- Background prediction
  - There is no data background prediction!
  - MC has stat and syst uncertainties (not shown here).



Background subtraction

$$\left(\frac{d\sigma}{dx}\right)_\alpha = \frac{\Sigma_j U_{j\alpha}(N_{data,j} - \boxed{N^{bkgd}_{data,j}})}{E_\alpha(\Phi T)(\Delta x)}$$

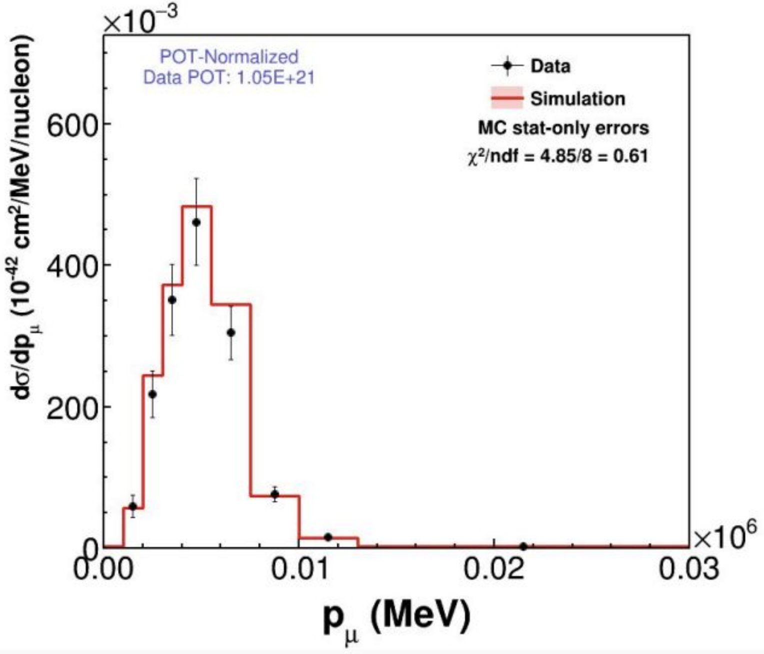
- This is the first place where you combine data and MC.
  - X = A - B
  - If A has error sources that B. doesn't have, X still inherits A's error sources.
- Data now has stat and syst errors.



Unfolding, efficiency, etc

- Next calculation steps follow the same idea, considering the errors.
- Your data now include the errors related to our use of imperfect simulation.

$$\left(\frac{d\sigma}{dx}\right)_{\alpha} = \frac{\sum_j \boxed{U_{j\alpha}}(N_{data,j} - N^{bkgd}_{data,j})}{\boxed{E_{\alpha}}\boxed{\Phi T}(\Delta x)}$$



Are we done yet?

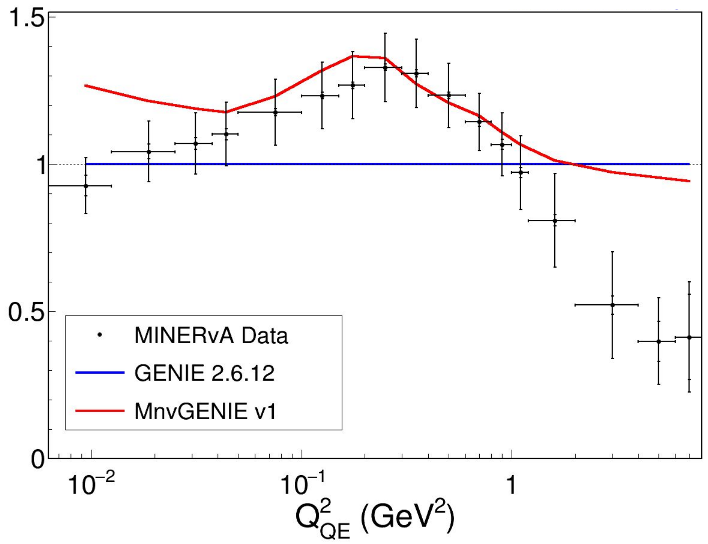
- No! We only talked about how do estimate the systematic uncertainties.
- There's a lot of work on trying to minimize and evaluate the uncertainties
  - E.g.: If the main source of a large uncertainty is known, the comparison can be reversed and, instead, a calibration can be obtained.
  - E.g.2: working on specific optimizations of your reconstruction (including the background channels) can reduce the errors.
  - ...

Model comparison

Neutrino
Interaction
School

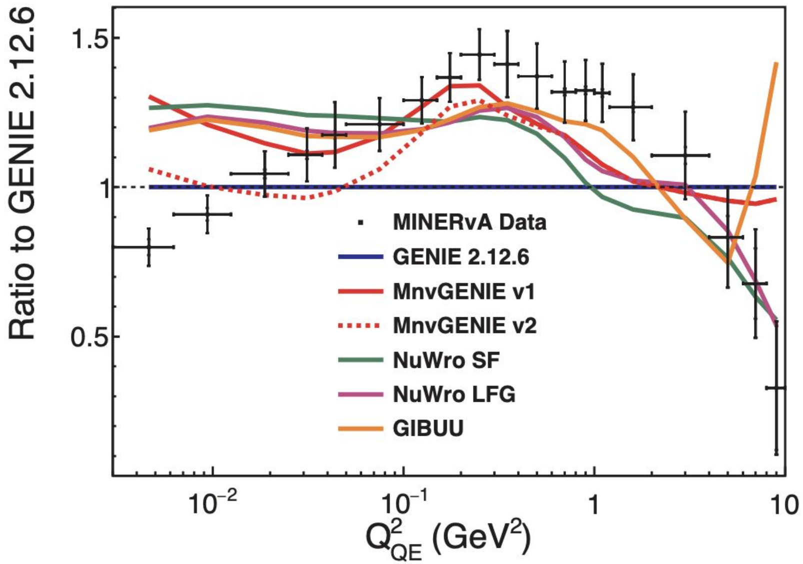
Model comparison

- Using the total uncertainty on the data allow us to compare it with simulated results.
- We can evaluate the regions where models describe well the data, and where they don't.



Model comparison

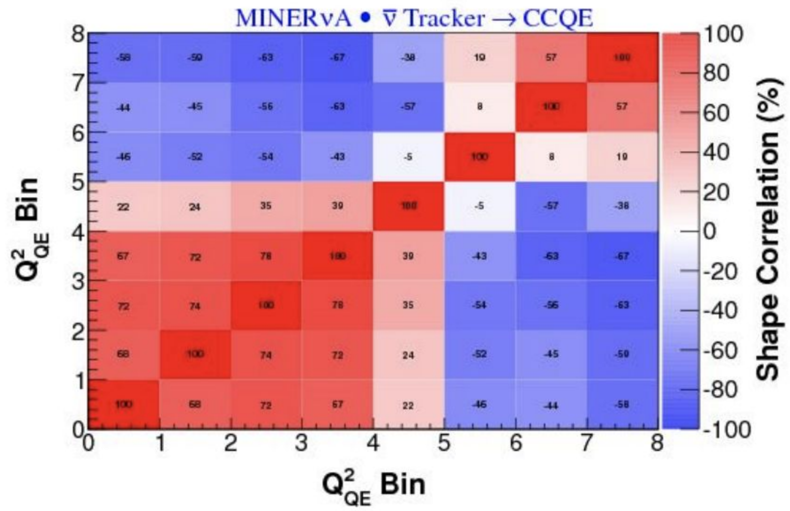
- Using the total uncertainty on the data allow us to compare it with simulated results.
- We can evaluate the regions where models describe well the data, and where they don't.



Covariance Matrix

- Although we already have the total uncertainty in each bin, we need to consider that different bins may be correlated.

- The covariance matrix is an N × N matrix, where N is the total number of bins in our measured distribution
  - Positive correlation
    - Universe shifts them in the same direction from the CV
  - Negative correlation
    - Universe shifts them in opposite directions from the CV



Goodness of fit

- There are many techniques to evaluate how good a model is, xsec experimental results often use chi-squared, where:
  - N is the number of bins
  - M_{ij}^{-1} is the $[i,j]$ matrix element of the inverse covariance matrix
  - x_i is the value of the ith bin of the quantity that the matrix was made from
  - y_i is the value of the ith bin of the model against which we are comparing

$$\chi^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - y_i)M_{ij}^{-1}(x_j - y_j)$$

Reporting results

- Hopefully it is somewhat clear that measurements need errors, and that those errors are intrinsic to the specific combination that was used to analyse the data

- Always make as much information publicly available as you can!
  - Variations used to evaluate errors
  - Values of unique errors
  - Covariance and Correlation matrices

- Always consider that someone in the future may way to re-evaluate your data and making that as easy as possible should be a priority.

Outro

We've gone through a lot!

- We went through the concept of stat vs syst errors, a bit of pdf's, we saw commons sources of systematics, learned how to estimate them, propagate the errors, used all of the information to compare data and monte carlo, and finally published it!
- This is not all, several of these steps can be done differently. But this talk should be enough for you to start considering errors in your first analysis.
- Uncertainties are unavoidable and they will be a big part of the job. It's complicated so don't feel bad if things are not quite clear yet. Trust me, even experts stumble with these concepts all the time. *Don't be shy and ask for help!*

We've gone through a lot!

- We went through the concept of stat vs syst errors, a bit of pdf's, we saw commons sources of systematics, learned how to estimate them, propagate the errors, used all of the information to compare data and monte carlo, and finally published it!
- This is not all, several of these steps can be done differently. But this talk should be enough for you to start considering errors in your first analysis.
- Uncertainties are unavoidable and they will be a big part of the job. It's complicated so don't feel bad if things are not quite clear yet. Trust me, even experts stumble with these concepts all the time. *Don't be shy and ask for help!*

Thank you!

Backup

Neutrino
Interaction
School

Backup

Statistics references

from Simon Connell at ASP2016 : Stats for HEP

G. Cowan, Statistical Data Analysis, Clarendon Press, Oxford, 1998.

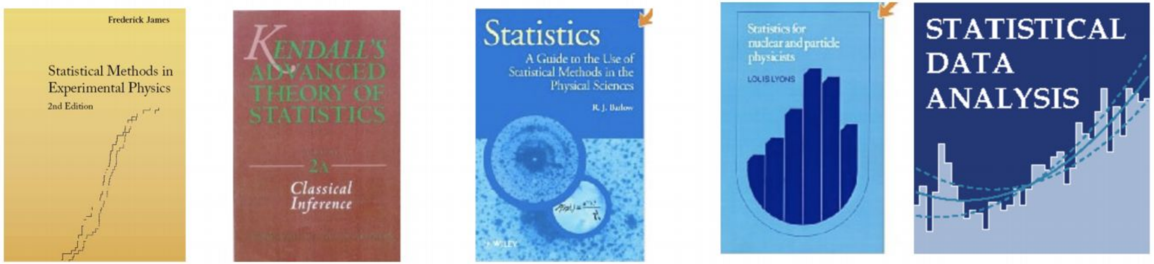
R.J.Barlow, A Guide to the Use of Statistical Methods in the Physical Sciences, John Wiley, 1989;

F. James, Statistical Methods in Experimental Physics, 2nd ed., World Scientific, 2006;

- W.T. Eadie et al., North-Holland, 1971 (1st ed., hard to find);

S.Brandt, Statistical and Computational Methods in Data Analysis, Springer, New York, 1998.

L.Lyons, Statistics for Nuclear and Particle Physics, CUP, 1986.



My favorite statistics book by a statistician:

Stuart, Ord, Arnold. "Kendall's Advanced Theory of Statistics" Vol. 2A *Classical Inference & the Linear Model*.

Statistical vs Systematic

E.g.: $x = 2.340 \pm$ **0.050 (stat.)** $\pm$ **0.025 (syst.)**

- Statistical or random uncertainties
  - can be reliably estimated by repeating measurements
  - follow a known distribution (e.g.. Poisson or a Gaussian) that can be measured by repetition
  - Relative uncertainty reduces as $1/\sqrt{n}$ where n is the sample size
  - Main HEP use case: Expect λ events in a search region, and observe n. The measurement error on λ is $\sqrt{n}$.

- Systematic uncertainties
  - Cannot be calculated solely from ‘sampling’ fluctuations (=repeated measurements)
  - In most cases don't reduce as $1/\sqrt{n}$ (but often also become smaller with larger n because more data allows better auxiliary measurements)
  - Difficult to determine, in general less well known than the statistical uncertainty. (HEP: typically >90% of the work)
  - Systematic uncertainties ≠ mistakes (a bug in your computer code is not a systematic uncertainty)

Backup

Covariance Matrix

For any two bins, i and j, the value of the covariance matrix element is given by

$$\boxed{M_{ij} = M_{ji} = \frac{\sum_{k=1}^{n}(x_{i_k} - \bar{x}_i)(x_{j_k} - \bar{x}_j)w_k}{\sum_{k=1}^{n} w_k}}$$

Where:

- M_{ij} is the matrix element corresponding to the covariance between bins i and j.

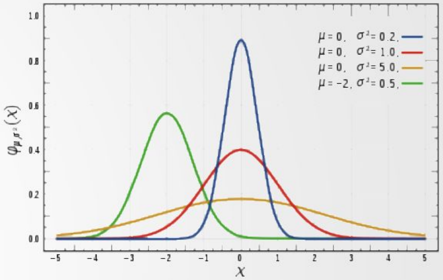
  Note that the covariance matrix is symmetric: $M_{ij} = M_{ji}$
- n is the total number of universes
- w_k is a weight to be applied to universe k (for unweighted universes, w_k is always 1)
- x_{i_k} refers to the event count in bin i, in universe k
- $\bar{x}_i$ is the mean event count in bin i, averaged over all universes: $\bar{x}_i = \frac{\sum_{k=1}^{n} x_{i_k} w_k}{\sum_{k=1}^{n} w_k}$

backup

Normal Distribution

- PDF: $g(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Mean: $\mathrm{E}(x) = \mu$
- Variance $V(x) = \sigma^2$

Important property:
If x_1, x_2 follow Normal distr. with μ_1,σ_1, and μ_2,σ_2 then x_1+x_2 follows Normal distr. with $\mu = \mu_1 + \mu_2$ and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.



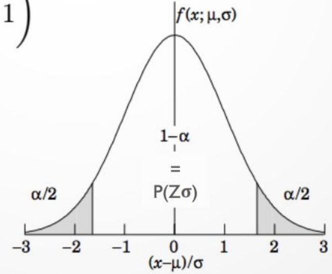
- "standard normal distribution" N(0,1): $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- Cumulative distribution of N(0,1):

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{d}z e^{-\frac{z^2}{2}} = \frac{1}{2}\left(\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)+1\right)$$

where 'erf' is the 'error' function.

$$P(Z\sigma) = \Phi(Z) - \Phi(-Z) = \mathrm{erf}\left(\frac{Z}{\sqrt{2}}\right)$$

- p-value: probability that a random process produces a measurement thus far, or further, from the true mean: $\alpha = 1 - P(Z\sigma)$



p-value α	δ
0.3173	1σ
4.55×10^{-2}	2σ
2.7×10^{-3}	3σ
6.3×10^{-5}	4σ
5.7×10^{-7}	5σ
2.0×10^{-9}	6σ

p-value α	δ
0.2	1.28σ
0.1	1.64σ
0.05	1.96σ
0.01	2.58σ
0.001	3.29σ
10^{-4}	3.89σ

Klaus Reygers, Heidelberg "Statistical Methods in Particle Physics", WS 2017/18

Poisson Distribution

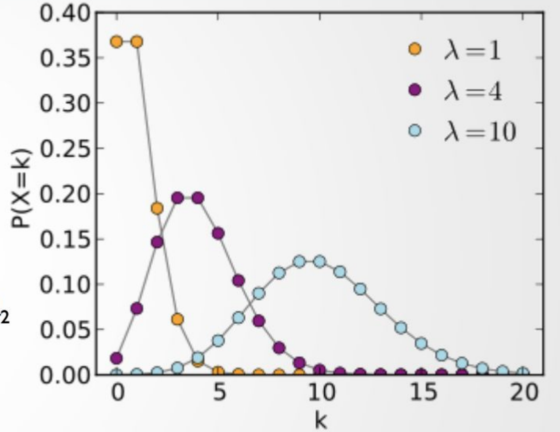
- Probability of a given number of events to occur in a fixed interval (of time or space) if these events occur with a known constant rate and independently of each other.

$$p(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$E(k) = \lambda$$

$$V(k) = \lambda$$

Important property:
If k_1, k_2 follow Poisson distr. with λ_1 , λ_2
→ k_1+k_2 follows Poisson distribution $\lambda_1+\lambda_2$



- Can be approximated by a Gaussian for large λ
- Examples:
  - Clicks of a Geiger counter in a given time interval
  - Number of Prussian cavalrymen killed by horse-kicks
  - Number of particle interactions of a certain type produced in a given time interval or for a given integrated luminosity

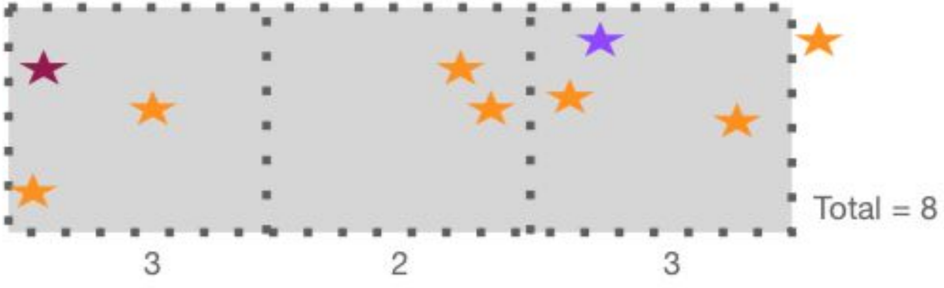
Number of deaths in 1 corps in 1 year	Actual number of such cases	Poisson prediction
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6

Klaus Reygers, Heidelberg "Statistical Methods in Particle Physics", WS 2017/18

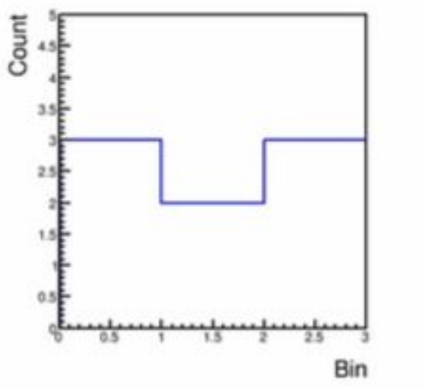
backup

Smearing shifts

Let's imagine we're trying to reconstruct how many events we detect in three areas of this "detector":



For our first universe - what happens if we shift the reconstructed event vertex to the left?



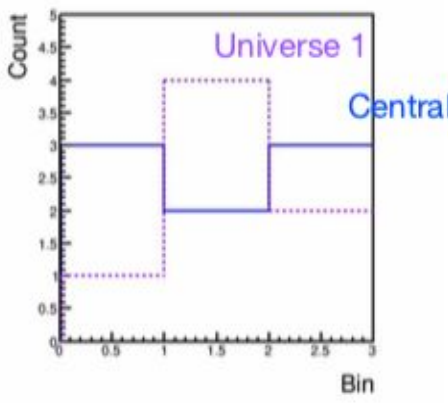
backup

Smearing shifts

This event has moved out of the sample

This event has moved to another bin

1 4 2 Total = 7



We call this a lateral shift because events move laterally from one bin to another - or even in and out of the sample

In this example the vertex position moves, but it could be a shift in energy, track angle…